

# Statistical Process Monitoring of Time-Dependent Data

Bart De Ketelaere\*, Tiago Rato<sup>†</sup>, Eric Schmitt<sup>‡</sup>, Mia Hubert<sup>§</sup>

## Abstract

During the last decades, we evolved from measuring few process variables at sparse intervals to a situation in which a multitude of variables are measured at high speed. This evidently provides opportunities for extracting more information from processes and to pinpoint out of control situations, but transforming the large datastreams into valuable information is still a challenging task. In this contribution we will focus on the analysis of time-dependent processes since this is the scenario most often encountered in practice, due to high sampling systems and the natural behavior of many real-life applications. The modelling and monitoring chal-

lenges that Statistical Process Monitoring (SPM) techniques face in this situation will be described and possible routes will be provided. Simulation results as well as a real-life dataset will be used throughout the paper.

**Keywords:** statistical process monitoring, time-dependent data, autocorrelation, nonstationarity, principal component analysis, cointegration

## 1 Introduction

Contemporary processes are typically highly automated, with in-line sensor technologies that produce vast amounts of data in a short period of time being the common situation. The result is the availability of large process streams that often display autocorrelation because of the fast sampling schemes relative to the process dynamics (i.e. inertial elements defining the settling time of the process). Additionally, in a substantial part of those real-life processes nonstationarity is an important factor. This scenario of multivariate, time-dependent data is one of the most challenging encountered in Statistical Process Monitoring (SPM), but it is often overlooked, although the separate fields of multivariate SPM and SPM for autocorrelated data have received more attention during the last decade (Woodall and Montgomery, 2014; Bersimis et al., 2007; Ferrer, 2007, 2014). In this contribution, research directions for monitoring such complex processes are highlighted.

The first direction is the latent variable approach. The idea is to project the data on a lower dimensionality, and to analyse these pro-

---

\*Bart De Ketelaere is research manager at the Division of Mechatronics, Biostatistics and Sensors (MeBioS) of the KU Leuven. His email is Bart.DeKetelaere@biw.kuleuven.be. He is the corresponding author. His affiliation is: Department of Biosystems, Division MeBioS, KU Leuven, Kasteelpark Arenberg 30, B-3001 Heverlee, Belgium.

<sup>†</sup>Tiago Rato is a researcher at the Division of Mechatronics, Biostatistics and Sensors (MeBioS) of the KU Leuven. His email is tiago.janeirorato@biw.kuleuven.be. His affiliation is: Department of Biosystems, Division MeBioS, KU Leuven, Kasteelpark Arenberg 30, B-3001 Heverlee, Belgium.

<sup>‡</sup>Eric Schmitt is a doctoral student in the Department of Mathematics, KU Leuven. His email is eric.schmitt@wis.kuleuven.be. His affiliation is: Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium.

<sup>§</sup>Mia Hubert is full professor in the Department of Mathematics, KU Leuven. Her email is mia.hubert@wis.kuleuven.be. Her affiliation is: Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium.

jected data (Kourti, 2005). This is especially rewarding when high cross-correlations are present so that large datasets can be projected on a low number of underlying variables. Such high correlations are increasingly common because of the use of spectral sensors (Near Infra Red (NIR) spectroscopy, Mass spectrometry, Frequency spectra, amongst others). In those situations as much as thousands of data points are collected for each sample. Principal Component Analysis (PCA) is a method that has shown success in coping with this setting and naturally leads to the Hotelling's  $T^2$  and the  $Q$  control charts that monitor the distance of a data point to the mean in the model space, and the distance between the original point and its projection, respectively. PCA and the derived control charts are not suitable for monitoring time-dependent data in their basic form, but extensions have been proposed in literature (e.g. (Ku et al., 1995; Wikstrom et al., 1998; Li et al., 2000; Wang et al., 2005; Kruger and Xie, 2012)) to address autocorrelation as well as nonstationarity. De Ketelaere et al. (2015) recently provided a literature review of different extensions to PCA that allow for monitoring time-dependent data. They divide those methods into non-adaptive and adaptive methods. The non-adaptive methods share the property that a model is built on an historical (calibration) dataset, and this model is then used to monitor data as they are acquired. Besides the classical PCA monitoring, also Dynamic PCA (DPCA) and Dynamic PCA with Decorrelated Residuals (DPCA-DR), due to Rato and Reis (2013b), belong to the non-adaptive methods. Because of the fact that the model parameters are not adapted throughout the monitoring process, these methods are typically more suited for monitoring processes where stationarity is assumed. When this assumption does not hold, it is better to adapt the model parameters to describe the new situation. Amongst the adaptive methods, Recursive PCA (RPCA) and Moving Window PCA (MWPCA) are the most known extensions to PCA (Li et al., 2000; Wang et al., 2005). In this paper, we

will discuss the use of PCA, DPCA, RPCA and MWPCA for time-dependent processes, and will focus mainly on the ability of these methods and the derived control charts to describe such data. We will also touch briefly upon a similar approach advocated in (Wikstrom et al., 1998), where the use of a classical PCA in combination with multivariate time series modeling of the scores is described.

A second direction comes from a completely different field, econometrics, and is to date largely unexplored in the SPM literature. In econometrics the situation of time-dependency and nonstationarity is omnipresent, and investigating the relation between different data series is essential, but is compromised by the fact that classical tests of association, e.g. based on the t-test, are not valid under nonstationarity (Phillips, 1986). Because of this, alternative tests are proposed, and approaches to model the dependency between multiple series are developed. A large part of this work is due to Granger and Newbold (1974) who developed the concept of cointegration which will be the basis of the monitoring approach that we will advocate. As far as the authors are aware of, using cointegration in an SPM setting was only mentioned in Chen et al. (2009) and De Ketelaere et al. (2011). Chen et al. (2009) conclude in their work that the cointegration testing method can be a useful methodology for engineering system condition monitoring and fault diagnosis, typically in systems under closed-loop control. De Ketelaere et al. (2011) also mention the potential merit of cointegration, but did not elaborate on this topic.

The goal of this discussion paper is thus to describe SPM methodologies for processes that are time-dependent. We will describe their basic working principles, apply them to typical datasets and discuss their strengths and weaknesses. Based on these, new directions of research will be proposed.

## 2 PCA-based methods for SPM of time-dependent processes

In this section we provide a brief overview of the modeling stages of typical PCA-based methodologies. A particular focus is given to the parametrization problems (selection of lags and forgetting parameters) and modeling assumptions. The impact of such factors on modeling performance will be then assessed in Section 2.2.

### 2.1 Algorithms

#### 2.1.1 Static PCA

Principal component analysis defines a linear relationship between the original variables of a dataset, mapping them to a set of uncorrelated variables.

In general, static PCA assumes that an  $(n \times p)$  data matrix  $\mathbf{X}_{n,p} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$  is observed. The sample mean of this dataset can be calculated as  $\bar{\mathbf{x}} = \frac{1}{n} \mathbf{X}_{n,p}' \mathbf{1}_n$  and its sample covariance matrix as  $\mathbf{S} = \frac{1}{n-1} (\mathbf{X}_{n,p} - \mathbf{1}_n \bar{\mathbf{x}}')' (\mathbf{X}_{n,p} - \mathbf{1}_n \bar{\mathbf{x}}')$ , where  $\mathbf{1}_n = [1, 1, \dots, 1]'$  is a vector of length  $n$ . PCA modeling proceeds by decomposing the sample covariance matrix as  $\mathbf{S} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}'$ , where  $\mathbf{P}$  is the  $p \times p$  loading matrix, containing columnwise the eigenvectors of  $\mathbf{S}$  and  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  has the respective eigenvalues in descending order. Afterwards, each  $p$ -dimensional vector  $\mathbf{x}$  is transformed into a score vector  $\mathbf{y} = \mathbf{P}'(\mathbf{x} - \bar{\mathbf{x}})$ .

In many cases, using  $k < p$  of the components still results in a good model. The  $k$ -dimensional scores are  $\mathbf{y}_k = \mathbf{P}_k'(\mathbf{x} - \bar{\mathbf{x}})$ , where  $\mathbf{P}_k$  contains only the first  $k$  columns of  $\mathbf{P}$ . Many methods exist to select the number of components to retain (see e.g. Valle et al. (1999) and Jolliffe (2002)). This study uses the Cumulative Percentage of Variance (CPV), which measures the amount of variation captured by the first  $k$  latent variables:

$$\text{CPV}(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} 100\%, \quad (1)$$

and  $k$  is selected such that the CPV is greater than a given threshold.

PCA control charts are based on the Hotelling's  $T^2$  statistic and the  $Q$  statistic (a.k.a. *Squared Prediction Error, SPE*). For any  $p$ -dimensional vector  $\mathbf{x}$ , the Hotelling's  $T^2$  is

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{P}_k \mathbf{\Lambda}_k^{-1} \mathbf{P}_k' (\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{y}_k' \mathbf{\Lambda}_k^{-1} \mathbf{y}_k \quad (2)$$

where  $\mathbf{\Lambda}_k = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$  is the diagonal matrix consisting of the  $k$  largest eigenvalues of  $\mathbf{S}$ . The  $Q$  statistic is defined as

$$Q = (\mathbf{x} - \bar{\mathbf{x}})' (\mathbf{I} - \mathbf{P}_k \mathbf{P}_k') (\mathbf{x} - \bar{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (3)$$

with  $\hat{\mathbf{x}} = \mathbf{P}_k \mathbf{P}_k' (\mathbf{x} - \bar{\mathbf{x}})$ .

If the number of observations is large, then assuming temporal independence and multivariate normality of the scores, the  $100(1 - \alpha)\%$  control limit for Hotelling's  $T^2$  is approximately the  $(1 - \alpha)$  percentile of the  $\chi^2$  distribution with  $k$  degrees of freedom; thus  $T_{\alpha}^2 \approx \chi_k^2(\alpha)$ .

A number of approximations exist to set the control limit of the  $Q$  statistic (e.g. Jackson and Mudholkar (1979), Box (1954)). We will use that of Box (1954), which shows that the  $Q$  statistic is approximately distributed as a scaled  $\chi^2$ -distribution with  $h$  degrees of freedom, denoted as  $g\chi_h^2$ . Provided that all eigenvalues of  $\mathbf{S}$  are available, the parameters are given by:

$$\theta_i = \sum_{j=k+1}^p \lambda_j^i \text{ for } i = 1, 2; \quad g = \frac{\theta_2}{\theta_1}; \quad \text{and } h = \frac{\theta_1^2}{\theta_2}. \quad (4)$$

The control limit for the  $Q$  statistic,  $Q_{\alpha}$ , is then taken as the  $(1 - \alpha)$  quantile of the  $g\chi_h^2$  distribution.

#### 2.1.2 Dynamic PCA

Dynamic PCA extends static PCA to autocorrelated, multivariate systems (Ku et al., 1995). DPCA works on the principle that in addition to the current observed variables, the respective lagged values up to a proper order,  $l$ , can also be

included in the PCA model. Therefore, DPCA applies PCA to an augmented dataset,  $\tilde{\mathbf{X}}(l)$ , constructed of lagged replicates of the original variables:

$$\tilde{\mathbf{X}}(l) = [\mathbf{X}(t), \mathbf{X}(t-1), \dots, \mathbf{X}(t-l)]. \quad (5)$$

Here  $\mathbf{X}(t-j)$  denotes the data matrix  $\mathbf{X}$  shifted  $j$  times into the past (i.e., with  $j$  lags). Due to its construction, DPCA implicitly fits an autoregressive model to the data. For instance, an AR(1) process will be modeled if lagged values up to order one are included in the model input, i.e.,  $\tilde{\mathbf{X}}(1) = [\mathbf{X}(t), \mathbf{X}(t-1)]$ .

Ku et al. (1995) provide an algorithm to specify the number of lags which adds an order to the lag structure, evaluates whether this brings any new linear relationship to the model and keeps it if it does. Experiments demonstrate that the number indicated by this method is often too low to model the process. More recently, Rato and Reis (2013a) detail an approach for selecting the number of lags by variable, allowing for a more refined model of the process being monitored. In this approach the appropriate number of lags is selected for each variable separately following a step forward procedure that explores the connection between small singular values and linear relationships of the data. In each stage, variables are tested, one at a time, for the inclusion of one lag and the lagged variable that leads to the smallest singular value is kept in the model. The procedure is repeated until a maximum number of lags is attained and the best combination of lags is then selected through an optimization function. Both lag selection algorithms were considered in this study, but result will only be discussed for the latter approach.

After building a DPCA model, the Hotelling's  $T^2$  and  $Q$  statistics are computed for an extended lagged vector that contains the current observation and its appropriate past values. The theoretical expressions for the control limits are analogous to those of static PCA and thus rely on the same i.i.d. assumptions.

### 2.1.3 Recursive PCA

If the stationarity assumptions of non-adaptive PCA models, such as those described above, are violated, then model parameter estimates obtained during the calibration phase may not be appropriate for future monitoring. Recursive PCA with a forgetting factor (Li et al., 2000) (RPCA) incorporates new observations and exponentially downweights old ones to update the mean and covariance matrix used in PCA.

Define the estimated mean and covariance of the observations up to time  $t$  as  $\bar{\mathbf{x}}_t$ , and  $\mathbf{S}_t$ . Then at time  $t+1$  the  $T^2$  and  $Q$  statistics are evaluated for the new observation  $\mathbf{x}_{t+1} = \mathbf{x}(t+1) = [\mathbf{x}_1(t+1), \dots, \mathbf{x}_p(t+1)]'$ . If both values do not exceed their cut-off value, the data matrix  $\mathbf{X}_{t,p}$  is augmented with observation  $\mathbf{x}_{t+1}$  as  $\mathbf{X}_{t+1,p} = [\mathbf{X}_{t,p}' \ \mathbf{x}_{t+1}]'$ . Next, the model parameters are updated by means of a forgetting factor  $0 \leq \eta \leq 1$ . Denoting  $n_t$  as the total number of observations measured until time  $t$ , the updated mean is defined as:

$$\bar{\mathbf{x}}_{t+1} = (1 - \frac{n_t}{n_t + 1}\eta)\mathbf{x}_{t+1} + \frac{n_t}{n_t + 1}\eta \bar{\mathbf{x}}_t, \quad (6)$$

and the updated covariance matrix is defined as:

$$\begin{aligned} \mathbf{S}_{t+1} = & (1 - \frac{n_t}{n_t + 1}\eta)(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})(\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1})' \\ & + \frac{n_t}{n_t + 1}\eta \mathbf{S}_t. \end{aligned} \quad (7)$$

This is equivalent to computing a weighted mean and covariance of  $\mathbf{X}_{t+1}$  where older values are downweighted exponentially. Using a forgetting factor  $\eta < 1$  allows RPCA to automatically give lower weight to older observations. As  $\eta \rightarrow 1$ , the model forgets older observations more slowly. The eigenvectors of  $\mathbf{S}_{t+1}$  are then used to obtain a loading matrix  $\mathbf{P}_{t+1}$ . Once the new value of  $k$  is determined (e.g. through a recalculation of the CPV) and the new eigenvalues calculated, the control limits of the Hotelling's  $T^2$  and  $Q$  statistics can be updated according to the formulas described earlier. Due to the fact that the model updates

throughout time for each new in-control point, it is essential that the method has a high power for detecting out-of-control (OOC) points. If not, those OOC points are included into the model updating scheme.

#### 2.1.4 Moving Window PCA

MWPCA updates at each time point while restricting the observations used in the estimations to those which fall within a specified *window* of time (Wang et al., 2005; Kruger and Xie, 2012). With each new observation, this window excludes the oldest observation and includes the observation from the previous time period. Thus, for window size  $H$ , the data matrix at time  $t$  is  $\mathbf{X}_t = [\mathbf{x}_{t-H+1}, \mathbf{x}_{t-H+2}, \dots, \mathbf{x}_t]'$ , and at time  $t+1$  it is  $\mathbf{X}_{t+1} = [\mathbf{x}_{t-H+2}, \mathbf{x}_{t-H+3}, \dots, \mathbf{x}_{t+1}]'$ . The updated  $\bar{\mathbf{x}}_{t+1}$  and  $\mathbf{S}_{t+1}$  can then be calculated using the observations in the new window. While completely recalculating the parameters for each new window is straightforward, and intuitively appealing, methods have been developed to improve on computational speed (see for example Wang et al. (2005) and Jyh-Cheng (2010)). As was the case for RPCA, the model is not updated when an observation is determined to be out of control and again the same control limits are used as described in the PCA section.

## 2.2 Simulation Studies

Typically, monitoring methods are evaluated for their fault detection in the literature, but good fault detection is predicated on a good model of the process and a correct definition of the related control limits. However, the monitoring approaches described in Section 2.1 do not ensure that an appropriate model is obtained for a broad range of process dynamics that are typical for real-life applications. Therefore, in this section we evaluate their validity through investigation of the modelling accuracy of the PCA-based methods on the AR(1) and ARI(1,1) processes. The AR(1) is cho-

sen as it is a widely encountered process dynamic in modern processes, and its integrated form (ARI) is used as its nonstationary counterpart. Following convention (e.g. Burnham et al. (1999); Choi et al. (2006)) we generate data at the subspace level so that we can explicitly control the features monitored by the PCA-based models. To obtain each observation at time  $t$  we began by generating five latent variables,  $\mathbf{y}_t$ , according to the equation of the desired AR(1) or ARI(1,1) process. For all processes, we introduce variation onto the process dynamics through  $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_5, 0.01\mathbf{I}_5)$ , where  $\mathbf{I}_5$  is the  $5 \times 5$  identity matrix. These are then transformed into a 50-dimensional dataset of measurements computed as

$$\mathbf{x}_t = \mathbf{P}_0 \mathbf{y}_t + \mathbf{e}_t, \quad (8)$$

where  $\mathbf{P}_0$  is a  $50 \times 5$  matrix with orthogonal columns randomly generated once and kept constant for all simulation runs. The  $\mathbf{e}_t$  are  $50 \times 1$  vectors of white noise errors, distributed as  $\mathcal{N}(\mathbf{0}_{50}, 0.000025\mathbf{I}_{50})$ , that simulate measurement noise, as is done, for instance, in (Ku et al. (1995) and Lakshminarayanan et al. (1997)). The  $\mathbf{e}_t$  can be seen as the error at the sensor level, and are set to a small value here under the assumption that sensors are typically reliable. For all methods and simulations, an arbitrary but common CPV of 95% is used.

The AR process is investigated because it is a particularly relevant process type seen the high sampling rate of many contemporary sensors inherently introducing (positive) autocorrelation into the data. Besides being a common process type in real life situations, AR processes have a natural relevance for studying the properties of DPCA.

The AR(1) process is defined as (Box et al. (1994)):

$$\mathbf{y}_t = \phi \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (9)$$

where  $\mathbf{y}_t$  are the serial observations of the underlying latent model ( $\mathbf{y}_t$  in Eq. (8)) and  $\phi$  is the AR coefficient. We consider values of  $\phi$  equal to 0,

0.1, 0.3, 0.5, 0.7, 0.9, with larger values of the parameter corresponding to stronger autocorrelation. Setting  $\phi = 0$  gives us a process with i.i.d. observations, which is the reference condition for which the assumptions of PCA and the theoretical control limits defined above are valid.

The ARI(1,1) process is defined as (Box et al. (1994)):

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \phi(\mathbf{y}_{t-1} - \mathbf{y}_{t-2}) + \boldsymbol{\varepsilon}_t. \quad (10)$$

Here the  $\phi$  values are the same as for the AR(1) case. When  $\phi = 0$  this process is simply an integrated process, I(1), or random walk. The ARI(1,1) process is considered because the adaptive methods were designed to address nonstationary processes.

Figure 1 depicts the AR(1) and ARI(1,1) processes with  $\phi$  values of 0.1 and 0.9. In general, when  $\phi$  increases, the variance of  $\mathbf{x}_t$  increases (with a factor  $(1-\phi^2)^{-1}$  in case of the AR(1)), so decreasing the relative effect of the noise. Also the unit root introduced in the nonstationary processes has a marked influence on the total signal variance, being even more pronounced than the effect of  $\phi$ . Because of both effects, different scaling factors were required in Figure 1 to visualize the typical behavior of the different simulation settings. Moving from left to right and top to bottom, the scaling factors used were 10, 5, 0.5 and 0.025. As a result, while the ARI(1,1) with  $\phi = 0.9$  appears to be relatively well behaved, its own scale is much greater than that of the other processes.

To assess the performance of the PCA-based models and corresponding control limits, 100 replicates of normal operation conditions (NOC) were generated. Each of these replicates is composed by 7000 NOC observations, divided into a calibration (first 6000 observations) and test (last 1000 observations) dataset. Models were specified for each replicate using the respective calibration dataset and their performances were subsequently assessed on the contiguous test dataset. False detection rates (FDR) were computed for each replicate as

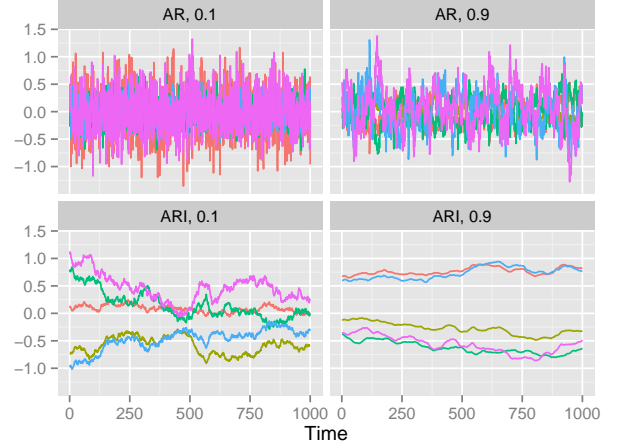


Figure 1: Plots of the rescaled AR(1) and ARI(1,1) processes with  $\phi$  values of 0.1 and 0.9.

the number of observations above the theoretical control limit divided by the total number of observations in the test phase. Therefore, for each process type and  $\phi$ , 100 FDRs were obtained. The distribution of the observed FDRs is then considered as a measurement of the models performance. For DPCA, RPCA and MWPCA additional parameters need to be chosen, such as the number of lags  $l$ , the forgetting parameter  $\eta$  and the window length  $H$ . In order to do so, an additional calibration dataset with 5000 NOC observations was generated for each combination of process type (AR(1) or ARI(1,1)) and  $\phi$  value. The number of lags used by the DPCA method was selected using the method of Rato and Reis (2013a).

Although the selection of the additional parameters for adaptive methods is critical to their proper implementation, this topic is not well covered in the literature. We based their choice on evaluating a range of possible values and assessing their appropriateness. The minimum and maximum values for  $\eta$  and  $H$  are  $[0.9, 0.9999]$  and  $[50, 2500]$ , respectively. For each process type models with these candidate parametrizations were applied to the additional calibration dataset, after splitting it into two equal parts. The first part is used to set up

the models with the given values of  $\eta$  and  $H$ , and the second part is then used to calculate the Sum of Squared Prediction Errors (SSPE) following the suggestions of Schmitt et al. (2014). The  $\eta$  and  $H$  values that minimizes the SSPE are then employed for process modelling. This approach can be thought of as a generalization of choosing the weighing factor in an EWMA control chart (Montgomery, 2008).

### 2.2.1 Simulation Results: AR(1)

For each of the simulation settings we considered, the parameter selection procedure explained above resulted in the parametrization which is given in Table 1. One trend that is apparent in this table is that the number of retained latent variables tends to decrease to the correct number, five, as  $\phi$  increases. The fact that the i.i.d. case did not lead to the underlying five latent variables is mainly due to the relatively large influence of the noise in these stationary cases and the chosen CPV value of 95%. The influence of the noise through  $\mathbf{e}_t$  is lowered when the autocorrelation increases (see Figure 1), explaining why, for higher values of  $\phi$ , the correct number of latent variables is extracted.

The impact of the dynamic features of the data is also visible on the lag selection procedures. In particular, the Ku et al. (1995) method selects zero lags for all AR(1) processes, except for  $\phi = 0.9$ , which has one lag (results not shown for the sake of brevity). Thus, this lag selection methodology finds that the dynamic relationships are not significant when the process exhibits moderate dynamics. This result is in line with the findings of Rato and Reis (2013a), who also concluded that the Ku et al. (1995) method has a tendency to underestimate the true dynamics of the data. As mentioned before, to overcome this issue, we present results for DPCA where the lag selection procedure of Rato and Reis (2013a) is implemented. In this approach the number of lags is not necessarily the same for all variables. For the cases study considered, the maximum number of lags was con-

sistently set as one, while the effective lag of each variable varied between zero and one. This means that some variables do not require any lag in order to describe the process data. Since DPCA also models the cross-correlation structure of the original as well as the lagged variables, the exclusion of redundant lags leads to more parsimonious models. It is noted, however, that in the i.i.d. case ( $\phi = 0$ ), the Rato and Reis (2013a) method on average adds one lag which is undesired. This happens because the optimization algorithm assesses the modelling improvements of consecutive lag structures and since the zero lag scenario is the first possible structure, it cannot be compared against a previous reference. Subsequently, the lowest feasible lag structure has at least one lag. Nevertheless, this situation can be avoided through further analysis of the data and decision graphs produced by the algorithm.

For the adaptive models, the selected forgetting parameters are all high, indicating that nonstationarity is not dominant in the data. Most of the values are at their upper bound, except for the large  $\phi$  cases. This does not come as a surprise since in case of large  $\phi$  the process mean does deviate from 0 for longer time periods, and the adaptive models try to capture these (random) dynamics by forgetting older observations faster.

Table 1: Parameter settings for monitoring methods in the AR(1) processes. Ranges are given for variable parameters, with the most frequent value in brackets.

$\phi$	PCA	DPCA		RPCA		MWPCA	
	$k$	$k$	Lags	$k$	$\eta$	$k$	H
0	7	14	0-1 (1)	7-8 (7)	0.9999	6-8 (7)	2500
0.1	8	14	0-1 (1)	6-8 (7)	0.9982	6-8 (7)	2500
0.3	6	12	0-1 (1)	6-7 (7)	0.9999	6-7 (7)	2469
0.5	5	10	0-1 (1)	5	0.9999	5-6 (5)	2496
0.7	5	10	0-1 (1)	5	0.9998	5	1775
0.9	5	7	0-1 (1)	5	0.9981	5	886

Monitoring was performed on each of the AR(1) settings and the false detection rates (FDR) of the Hotelling's  $T^2$  and  $Q$  statistics were recorded. The desired overall FDR is set at 1%, and since we

have no knowledge about the correlation between the  $T^2$  and  $Q$  statistic, it is assumed to be zero and the Bonferroni correction is applied such that  $\alpha_{T^2} = \alpha_Q = FDR/2 = 0.005$ . Boxplots of the FDRs for the Hotelling's  $T^2$  and the  $Q$  statistics are presented in Figures 2-5, as a function of the autocorrelation parameter  $\phi$ .

Across the results, we see that the effect of autocorrelation on the modelling properties of the PCA-based methodologies is not strong, except for high values of  $\phi$ . This effect has a greater influence in the Hotelling's  $T^2$  statistic dynamics since the original autocorrelation of the data is directly translated to the scores, which ultimately compromises the reliability of theoretical control limits (Kruger et al., 2004; Vanhatalo and Kulachi, 2015). Although the observed false detection rate of the Hotelling's  $T^2$  statistics is generally within expectation, we see that the dispersion of the FDR values increases as the autocorrelation increases. This is a direct result of the inherent dynamics on the Hotelling's  $T^2$  statistics, since it increases the probability of having consecutive measurements with similar values. Thus, for replicates where the process experiences sustained deviations from the model (i.e., high values of Hotelling's  $T^2$ ), the false detection rate is higher than specified, while the converse happens when the process runs close to the model. In that case, the Hotelling's  $T^2$  statistic exhibits consecutive, low values. This results in more variable detection performance, even though the average FDR is close to the desired value. Although the FDRs obtained for the  $T^2$  are generally in line with expectations, extensions to the PCA framework can produce some additional improvement. This was the idea behind the approach of Wikstrom et al. (1998), which applies an ARIMA modelling approach to the scores. We indeed observed a modest decrease in this dispersion for high values of  $\phi$ , but since the Wikström approach does not consider the residual space, it cannot solve the problem seen with the  $Q$  statistics.

On the other hand, since the  $Q$  statistic is re-

lated with the model residuals, it should be serially decorrelated as long as the appropriate number of latent variables is retained. This in turn should lead to good monitoring performance. We observe that this is the case for low values of  $\phi$ , since reasonable FDR are obtained. For larger values of  $\phi$  the models also produced serial decorrelated residuals. However, the scores subspace is not accurately explaining the dynamic characteristics of the data, causing the residuals to be greater than expected, which leads to a higher FDR than the target. While extensions based on classical time series methods, such as that of (Wikstrom et al., 1998), are applicable to the  $T^2$  statistic as we mentioned above, the number of variables used to calculate the  $Q$ -statistic can be extremely large, and therefore beyond the capacity of such methods.

By using the correct number of lags in the DPCA model, an elimination or at least reduction of some of the misspecifications is expected, resulting in an improved modelling and monitoring performance. However, even though the DPCA approach is able to follow the process dynamics more closely, it still produces monitoring statistics (especially the Hotelling's  $T^2$ ) with dynamic characteristics. This indicates that DPCA is prone to the same deficiencies identified earlier for PCA, which essentially lead to misspecified control limits.

The adaptive methods show similar results as the non-adaptive PCA and DPCA methods. The reason for this are the very large forgetting parameters  $\eta$  and window size  $H$  presented in Table 1. These cause the methods to forget only very slowly. These large values indicate that nonstationarity is not a major issue for this process, which in fact is correct. For MWPCA, in case of  $\phi = 0.9$ , the window size  $H$  was substantially smaller, causing the FDR dispersion to be substantially larger for the Hotelling's  $T^2$  statistic.

## 2.2.2 Simulation Results: ARI(1,1)

Next, we consider the ARI(1,1) process, again setting the target FDR equal to 1%. The ARI(1,1)



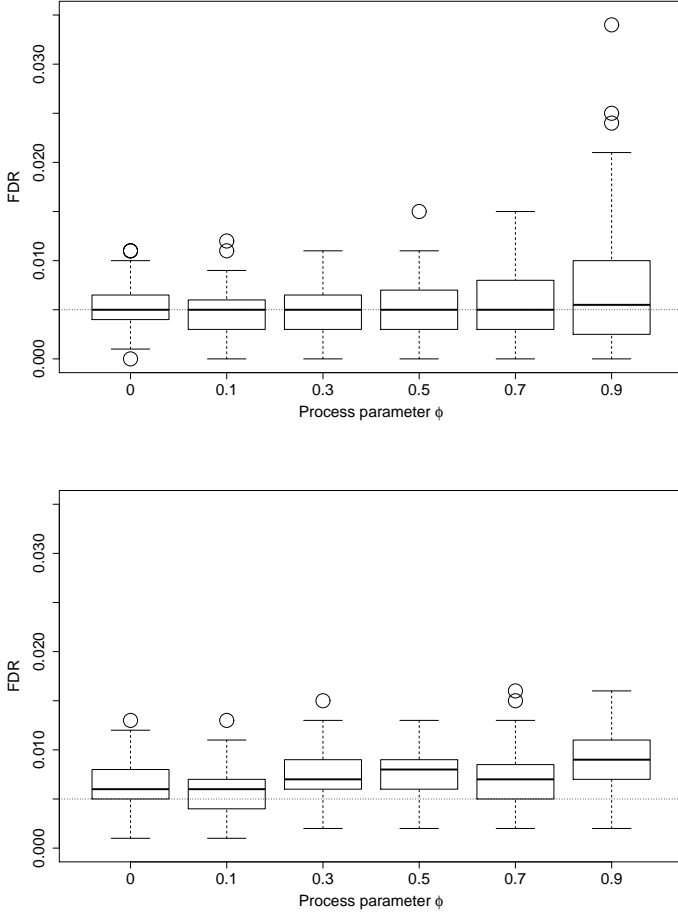


Figure 2: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of PCA on the AR(1) process with  $\phi$  ranging from 0 to 0.9. The value of  $\alpha_{T^2} = \alpha_Q = 0.005$ .

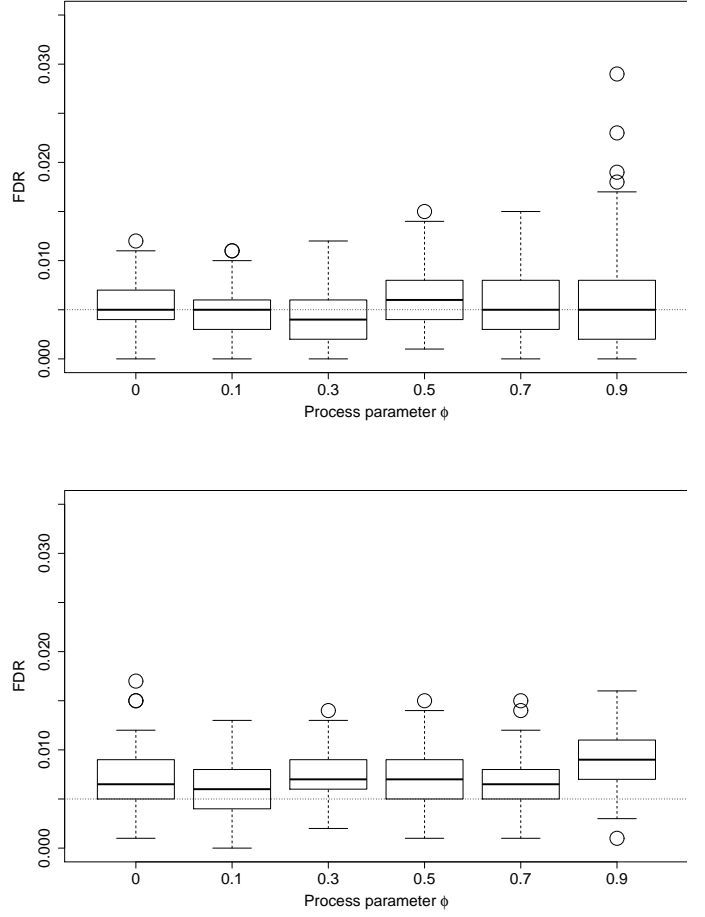


Figure 3: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of DPCA on the AR(1) process with  $\phi$  ranging from 0 to 0.9. The value of  $\alpha_{T^2} = \alpha_Q = 0.005$ .

process poses a greater monitoring challenge for PCA and DPCA than the AR(1) process because of the apparent nonstationarity (see Figure 1). This is expected since neither of these methods is designed to cope with nonstationary behavior. Experiments confirmed that FDRs typically reach 100% for these methods regardless of the value of  $\phi$ , so results are not shown. This is caused by the fact that the data used to build the models are not representative for new data encountered in the test

dataset. The approach of Wikström et al. (1998) does fail as well when an ARMA model is fitted through the  $T^2$  statistics because they are also nonstationary, so that differencing the scores is required. Furthermore, the nonstationarity around the PCA model remains unexplained by the Wikström approach.

The model parametrizations of the adaptive methods are shown in Table 2. We can see that in general, as  $\phi$  increases, the forgetting factor de-

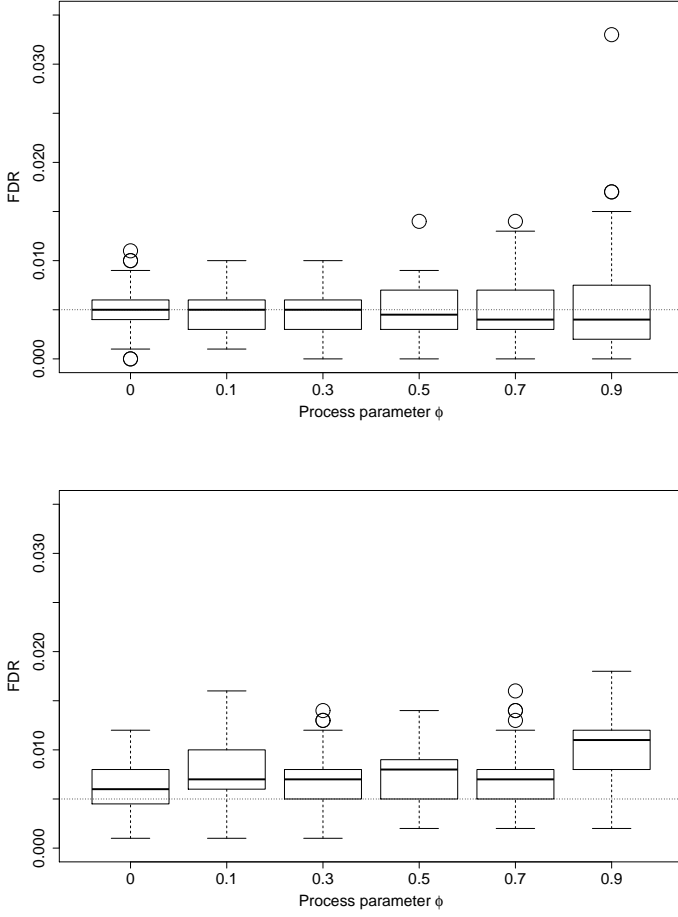


Figure 4: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of RPCA on the AR(1) process with  $\phi$  ranging from 0 to 0.9. The value of  $\alpha_{T^2} = \alpha_Q = 0.005$ .

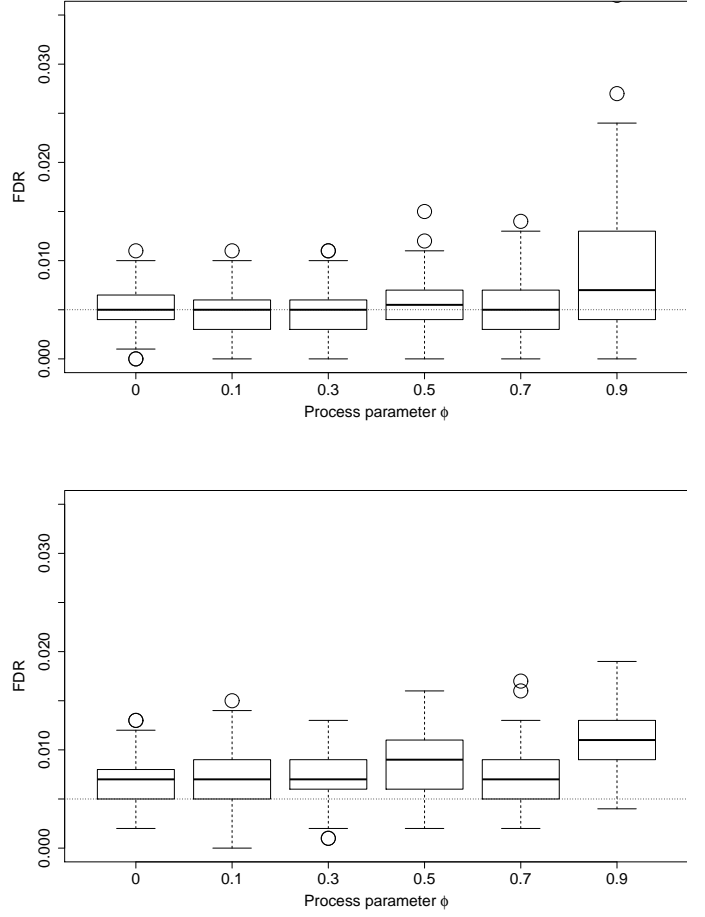


Figure 5: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of MWPCA on the AR(1) process with  $\phi$  ranging from 0 to 0.9. The value of  $\alpha_{T^2} = \alpha_Q = 0.005$ .

creases, although there are deviations from this pattern. This was to be expected since those situations are dominated by the strongest nonstationarity as we have demonstrated in Figure 1, and is in line with the observation for the AR(1) case. RPCA and MWPCA are expected to perform acceptably in this setting since they are able to adapt to process changes. However, both methods produce unacceptable results across all values of  $\phi$  when the theoretical control limits are used in com-

bination with  $\alpha = 0.01$ . Changing the forgetting factors to improve results did not lead to consistently on-target performance.

The reason for this poor behavior is two-fold. First, when applying RPCA and MWPCA, the models are only updated when a new point is considered in control. When the forgetting factor is not chosen ideally, or when the dynamics of the underlying process change, the adaptive methods can fail to follow those dynamics, leading the model

Table 2: Parameter settings for monitoring methods in the ARI(1,1) processes. Ranges are given for variable parameters, with the most frequent value in brackets.

$\phi$	RPCA		MWPCA	
	$k$	$\eta$	$k$	H
0	3-5 (4)	0.9981	3-5 (4)	2500
0.1	3-5 (4)	0.9986	3-5 (4)	1400
0.3	3-5 (4)	0.9972	4-5 (4)	700
0.5	3-5 (4)	0.9955	3-5 (4)	450
0.7	3-5 (4)	0.9800	3-5 (4)	250
0.9	3-4 (4)	0.9500	2-4 (3)	100

to consider a large portion of the data to be out of control. As stressed before, the right choice of the forgetting factor and its eventual updating to account for changing dynamics is important, but references are scarce (e.g. Choi et al. (2006)) and the topic deserves further attention.

The second reason for the excessive FDR comes from the fact that the underlying assumptions of the analytical Hotelling’s  $T^2$  and  $Q$  limits defined earlier and applied here with  $\alpha$  set at 1% do not hold. This misspecification causes the control limits to be too tight, so that a substantial number of observations are considered outlying. This in turn prevents the model from being updated since such OOC points are not used for adapting the model. As advocated in (Rato et al., 2015), in such cases it is better to tune the  $\alpha$  value such that the desired FDR is obtained (for other examples of approaches for adjusting the limits, see e.g. (Ramaker et al., 2005), (Camacho et al., 2009), and (van Sprang et al., 2002)). In Figures 6 and 7, we consider smaller values of  $\alpha_{T^2}$  and  $\alpha_Q$  (see Table 3), resulting in higher control limits but acceptable FDR rates. These  $\alpha$  values were determined manually by dividing a fixed reference data set in two parts, fitting a model to the first part (5000 observations), and assessing its performance on the second part (15000 observations). Then, the selected  $\alpha$  values were applied to all 100 simulation runs. The variable performance in the simulations shows that this approach, while generally effective, does not result in models that generalize

to all of the realizations of the process encountered in the simulations. The figures demonstrate that for a substantial number of the simulation runs the FDR is far from the target FDR, with values that reach 100%. The reason for those extreme cases is that models are only updated when a new point is considered in control. If at the start of the monitoring phase points are considered OOC the model does not update, increasing the probability that later measurements will be considered OOC as well when the process further deviates from the model because of the nonstationarity. Table 3 lists the median FDR values since they are not clearly visible in the boxplots. For the  $T^2$  values, the median is actually zero, meaning that no NOC points were considered to be outlying. This illustrates that tuning  $\alpha$  for the Hotelling’s  $T^2$  statistic is not working adequately and improvements are needed. This is also visualized in Figure 6, where indeed the  $T^2$  control limit is too high because of the substantial autocorrelation present in the Hotelling’s  $T^2$  statistics.

For the  $Q$  statistic, the medians are in line with expectations, meaning that if the method is actively monitoring, the tuned control limit for the  $Q$  is adequate. This is illustrated in Figure 8, where the  $Q$  statistic shows a random behavior in periods of similar limits, and where the amount of adaptation of the control limit in periods of higher/lower residuals is acceptable.

As mentioned above, the tuned  $\alpha$  values in Table 3 were selected by trial an error adjustment on a validation dataset because no better approach is currently available. From these results we can observe a direct relationship between the value of  $\phi$  and the values of  $\alpha$  needed to have a FDR that is in line with expectations: the higher  $\phi$ , the smaller the  $\alpha$  values must be. The cause of this relationship is that increasing  $\phi$  increases the process variance, and this increased process variance is not accounted for in the theoretical limits. Therefore, for processes with only moderate nonstationarity, the problems we observe in this simulation may not arise to the same extent, and results listed here can

be considered as extremes (worst case scenario). We note that the interquartile range does not show a clear trend except that it is typically larger for the  $T^2$ -statistic than for the  $Q$ -statistic. It might be that even further refined model parametrizations or additional simulation runs are required for a more obvious pattern to emerge.

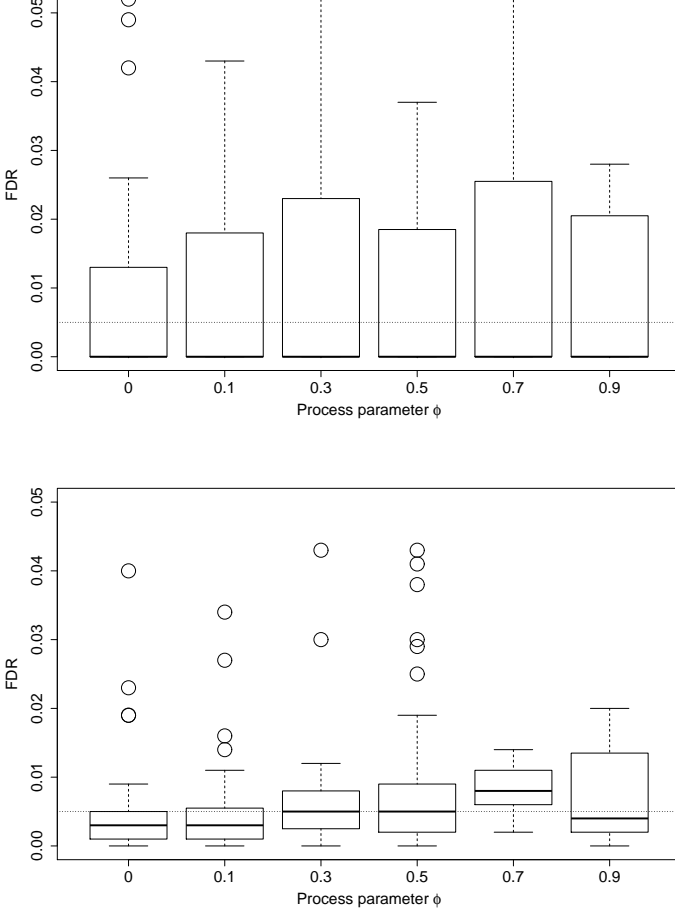


Figure 6: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of RPCA on the ARI(1,1) process with  $\phi$  ranging from 0 to 0.9, using the tuned values of  $\alpha$  in Table 3.

A visual appreciation of the monitoring behavior of RPCA applied to an ARI(1,1) process with  $\phi = 0.9$  and tuned  $\alpha$  values is given in Figure 8.

Table 3: Tuned  $\alpha$  values for monitoring methods in the ARI(1,1) processes with an intended overall FDR of 0.01 ( $FDR_{T^2} = FDR_Q = 0.005$ ). The observed FDRs are summarized by their median and interquartile range.

$\phi$	Stat.	RPCA		MWPCA	
		$\alpha$	FDR	$\alpha$	FDR
0	$T^2$	$2 \times 10^{-2}$	0 (0.013)	$5 \times 10^{-4}$	0 (0.012)
	$Q$	$10^{-3}$	0.003 (0.004)	$5 \times 10^{-4}$	0.002 (0.019)
0.1	$T^2$	$2 \times 10^{-2}$	0 (0.018)	$1.5 \times 10^{-4}$	0 (0.011)
	$Q$	$1.5 \times 10^{-3}$	0.003 (0.005)	$10^{-4}$	0.003 (0.017)
0.3	$T^2$	$2 \times 10^{-2}$	0 (0.023)	$5 \times 10^{-5}$	0 (0.009)
	$Q$	$1.5 \times 10^{-3}$	0.005 (0.007)	$5 \times 10^{-5}$	0.003 (0.013)
0.5	$T^2$	$1.5 \times 10^{-2}$	0 (0.019)	$6 \times 10^{-5}$	0 (0.006)
	$Q$	$10^{-3}$	0.005 (0.007)	$10^{-5}$	0.003 (0.009)
0.7	$T^2$	$1.2 \times 10^{-2}$	0 (0.026)	$2 \times 10^{-5}$	0 (0.012)
	$Q$	$1.5 \times 10^{-5}$	0.008 (0.005)	$10^{-7}$	0.004 (0.003)
0.9	$T^2$	$1.7 \times 10^{-2}$	0 (0.021)	$1.5 \times 10^{-6}$	0 (0.018)
	$Q$	$5 \times 10^{-6}$	0.004 (0.012)	$10^{-9}$	0.006 (0.005)

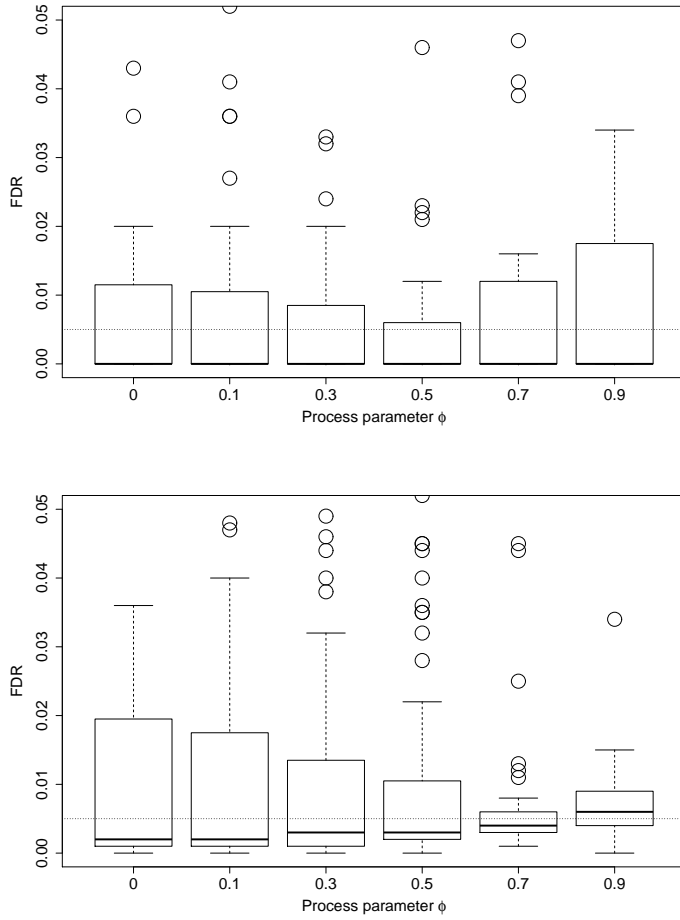


Figure 7: False detection rates of the  $T^2$  (top) and  $Q$  statistics (bottom) of MWPCA on the ARI(1,1) process with  $\phi$  ranging from 0 to 0.9, using the tuned values of  $\alpha$  in Table 3.

From this figure we conclude that the monitoring statistics still show evidence that the model is not completely explaining the structure of the process, which is especially visible in the  $T^2$  statistic. In fact, the recursive nature of RPCA does allow to cover the simple case of nonstationarity, but does not seem to cope with the AR component that is added to it.

Even though this parametrization reduces the detection of small faults, the model is adapting

and actively monitoring. Therefore, if the process of interest displays large faults, these methods may still be suitable for monitoring.

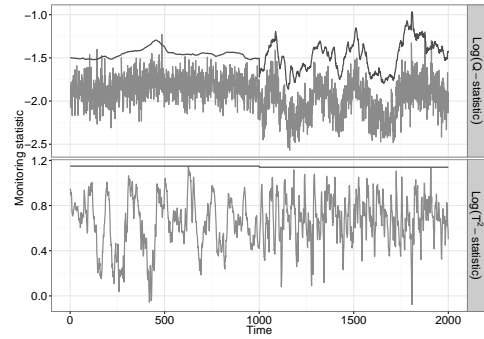


Figure 8: RPCA-based control chart for the ARI(1,1) process with  $\phi$  equal to 0.9. The value of  $\alpha_{T^2} = 1.7 \times 10^{-2}$  and  $\alpha_Q = 5 \times 10^{-6}$ . The control statistics are in  $\log_{10}$  to improve readability.

## 2.3 Discussion and future perspectives

Based on simulation results we covered different forms of time-dependency in process monitoring, focusing on the simple yet challenging cases of an AR(1) and an ARI(1,1) because those types of dynamics are believed to be often present in modern process data.

The results of the AR(1) simulations demonstrate that under moderate dynamics, all of the studied PCA-based methodologies have a similar, acceptable modelling performance. This happens because the optimal parameters of DPCA, RPCA and MWPCA tend to reduce them to static PCA. It is also visible that when process dynamics become more relevant (say, for  $\phi \geq 0.7$ ) the models tend to deviate more from expectation, with especially the Hotelling's  $T^2$  statistic to be less reliable, confirming recent results from (Vanhatalo and Kulachi, 2015). However, simple AR(1) dynamics do not severely compromise the modelling capabilities

of the procedures, which are still producing false detection rates within expectation.

On the other hand, the ARI(1,1) simulations showed that PCA and DPCA cannot cope with nonstationarity. RPCA and MWPCA, the adaptive methods that are devised for handling nonstationarity, do allow for modeling such data, but plugging the classical values for  $\alpha$  in the control limits for the Hotelling's  $T^2$  and  $Q$  statistic resulted in FDR values that were unacceptably high. Since no literature is available for defining those control limits under the nonstationarity assumption, it was proposed to relax the control limits for both statistics by searching for  $\alpha$  values that result in acceptable FDRs so that these models could continue to adapt to the time-varying process and might be able to detect severe faults. The observation that there is a clear link between the process dynamics and the monitoring method capable of handling such data is partly in line with the results of Camacho et al. (2009), who acknowledge that it is important to reflect the time-varying nature of the process in the model of the SPM method used. Interestingly, Camacho et al. (2009) mention the fact that besides the process dynamics also the fault type to be detected is important when deciding on the best monitoring method. This is ultimately true, and fault detection is probably the most important aspect when considering SPM methods. As an example, (Rato et al., 2015) concluded that the capability of the adaptive methods to detect ramp faults is highly dependent on the forgetting factor chosen, and should be considered carefully.

The nonstationarity as introduced into the simulations are extreme cases, as can be seen from Figure 1. Performance may be more appropriate on processes with less severe forms of nonstationarity, like processes showing mild nonstationarity or simple drifts due to sensor aging. In those cases, the proposed extensions to PCA might be able to capture the drifts thus describing the data adequately. Contrarily, in order to turn these models into valid and powerful monitoring schemes work is required

into a proper definition of the control limits connected to the methods. We believe that this direction of research is highly relevant. In case of more complex nonstationary behavior, work is required into the modeling aspect as we demonstrated in the ARI(1,1) case and high  $\phi$  values where even after adapting  $\alpha$  deviating FDR values were noted.

Although nonstationarity is present in a wide range of processes, the use of adaptive models is still limited, especially in the multivariate case. The moderate results we have shown are only a partly explanation. From the practice side, the lack of intuition with the methods by the process owners themselves (often engineers) is a barrier as well. From that perspective, it could be advantageous to translate the parameter choice of e.g. the RPCA into a selection procedure for parameters engineers are used to. More specifically, engineers typically have a good idea of the process dynamics in terms of their in control *frequency spectrum*, i.e. the speed of change which is typical to those processes. This behavior can be visualized through the generation of the Power Spectral Density (PSD) of the process, denoting the power of the signal as a function of the frequency (Oppenheim and Schaffer, 1975). Typically, only the slow dynamics are proper to the underlying process, so that the SPM scheme should apply a low pass filtering of the data. In essence, the cut-off frequency determining the frequencies which do (not) pass the SPM model are well related to the AR and/or MA terms. Said this, we feel that bridging the gap between the engineering and statistical reasoning could help implementation of adaptive SPM methods. This implementation issue is the last barrier: software to cope with such multivariate SPM models is not wide-spread, and is required to translate advanced multivariate SPM from a pure research field into a practical solution.

### 3 Cointegration and Error Correction Models for SPM

Above, we discussed extensions to PCA as plausible solutions to handle multivariate data series showing time-dependent behavior. Besides this obvious choice, we will present a different, general approach which has not been fully explored in an SPM setting. The approach we will advocate has its roots in econometrics, a discipline where non-stationary time series are a frequent issue.

#### 3.1 Illustrating examples

Granger, a Nobel Prize winner for his work on non-stationary time series and causality, showed that when stationarity cannot be assumed, correlations amongst variables are often spurious ("spurious regressions") and the asymptotic behavior of classical tests do not hold (Granger and Newbold, 1974). This dictates the need for alternatives. The behavior of spurious correlations are demonstrated using 1000 realizations of a simple i.i.d. process that consists of 1000 data points, versus 1000 realizations of a random walk of 1000 data points, which is actually the integration of the i.i.d. process. We will denote the i.i.d. process by  $I(0)$ , and the random walk by  $I(1)$ , since it is a first order integrated (I) process having no autocorrelation (AR) nor moving average (MA) term. Both processes are visualized in Figure 9.

Evidently, both the i.i.d. processes as well as the random walks are purely driven by randomness, and any correlation between the different realizations is random. However, when we plot the sample correlations between the  $I(0)$  series on the one side and the  $I(1)$  series on the other side, we see that they behave very differently (Figure 10). For the i.i.d. case, 5.5% of the correlations are significant at  $\alpha = 0.05$ , completely within expectations. For the random walks this percentage is much higher, i.e., 91% of the correlations are

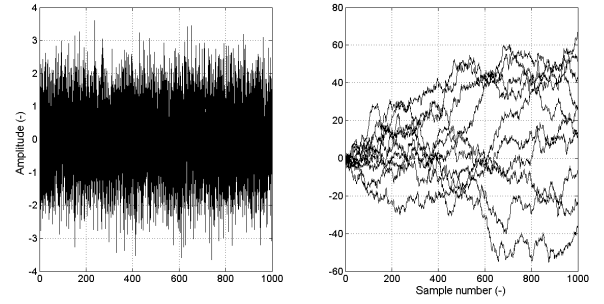


Figure 9: 1000 white noise (i.i.d.) sequences of length 1000 (left); 1000 random walk sequences of length 1000 (right).

significantly different from zero. This simple example shows that classical statistical tests do not hold in the presence of nonstationarity, and alternatives are needed. Granger and co-workers developed tests to check whether the correlation between nonstationary series is spurious or not. If it is so, no statistical analysis is meaningful, but if the series are causally related, several techniques are proposed to analyze their relation under the assumption of nonstationarity. This last point is interesting because in reality we most often acquire signals which all relate to different characteristics of a given process, so that they are expected to be causally related (namely, through the underlying process).

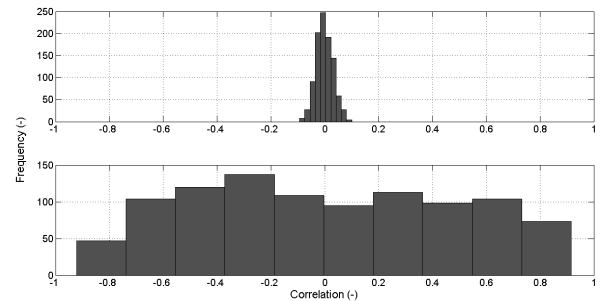


Figure 10: Pearson correlation coefficients for the  $I(0)$  (top) and the  $I(1)$  (bottom) situation.

An example of such nonstationary series from an industrial example is depicted in Figure 11 where seven temperature profiles are shown. Those temperatures are used to monitor the bearings of a rotating machine. Because of variable load of the machine, fluctuating ambient temperature and wind speed, the temperatures vary widely as function of time. The large drop nearly half-way the series is related to a stand-still and restart of the machine. Given those fluctuations which are observable but uncontrollable and unpredictable, it is expected that temperature will rise when problems occur with the bearings. This is typically seen as short increases due to temporal blocking, and those are most often much smaller than the observed temperature variation under normal operation. Such data are challenging because of the nonstationarity, and the fact that they do not allow to simply subtract an average trajectory as is often done in batch processes.

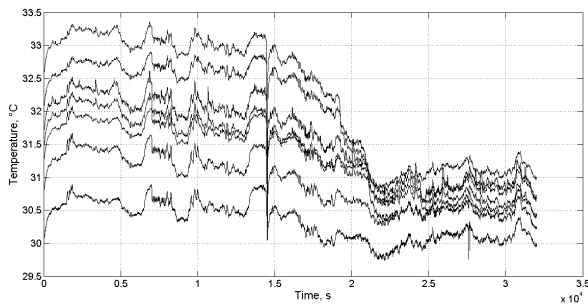


Figure 11: Machine temperature change over time at 7 locations.

In the following paragraph, we will briefly introduce tests for (Granger) causality and the basic form of cointegration, a framework to analyse the relation between nonstationary series. We will use the practical case of temperature profiles to show the merits of this approach. As it is only the goal to set the scene without going into depth, we will explain the general methodology based on two temperature series, and will provide guidelines for generalization later on.

### 3.2 Cointegration basics

Suppose we have a process for which two time series are acquired,  $\mathbf{x}_t$  and  $\mathbf{y}_t$ , and suppose they are integrated of order 1,  $I(1)$ . The regression of  $\mathbf{y}_t$  on  $\mathbf{x}_t$ ,

$$\mathbf{y}_t = c_0 + c_1 \mathbf{x}_t + \varepsilon_t, \quad (11)$$

will yield high correlations, but as we have discussed before those might be spurious. If  $\varepsilon_t$  is also  $I(1)$ , the OLS estimators become inconsistent (Phillips, 1986). However, when the above regression yields stationary residuals, the OLS estimators are consistent, and we say that  $\mathbf{x}_t$  and  $\mathbf{y}_t$  are cointegrated and further analysis makes sense. Formally, we want to test

$$\begin{aligned} H_0 : \mathbf{x}_t \text{ and } \mathbf{y}_t \text{ are not cointegrated,} \\ H_1 : \mathbf{x}_t \text{ and } \mathbf{y}_t \text{ are cointegrated.} \end{aligned} \quad (12)$$

Under  $H_0$ , the absence of cointegration between  $\mathbf{x}_t$  and  $\mathbf{y}_t$  is assumed, so if we reject it we have evidence of cointegration. Only in that case estimation of the relation given in Eq. 11 makes sense. This relation is called the equilibrium, or long run relation (Engle and Granger, 1987). The approach of fitting the equilibrium relation and checking for stationarity of the residuals is often referred to as the Engle-Granger approach (Engle and Granger, 1987). The check of the absence of a unit root in the residuals can be performed using the Augmented Dickey Fuller test (ADF) (Fuller, 1976; Phillips, 1987). The ADF is illustrated for two random walks (Figure 12) from the earlier example displaying a high correlation ( $r = 0.74$ ). Applying the ADF on the residuals of the regression between both random walks results in not rejecting the null hypothesis ( $p > 0.05$ ), and thus to the conclusion that the series are not cointegrated.

If we now take two temperature profiles from the industrial dataset, which clearly display nonstationarity, we observe a high correlation as well ( $r = 0.92$ ). However, performing the ADF test



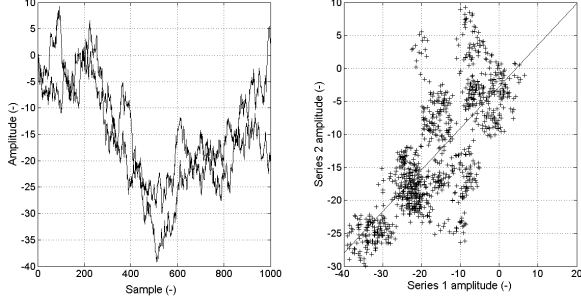


Figure 12: Example of two random walks (left) displaying a high correlation and the regression of Series 2 on Series 1 (right).

rejects the null hypothesis ( $p < 0.05$ ), implying that both series are cointegrated, and that the long term relation between both series makes sense. This relation is visualized in Figure 13 and although it succeeds in describing the global relation, at a shorter time span the series seem to diverge and converge again. In fact, the two series  $\mathbf{x}_t$  and  $\mathbf{y}_t$  being cointegrated means that they are attracted towards the long term equilibrium given in Eq. 11 (Maddala and Kim, 2003). The long term relationship reflects the general behavior of the machine temperature as function of load and ambient temperature ("steady state"), whereas the short term behavior is typical for the transient behavior. When the load of the machine is altered, the temperature will start changing, but this change depends amongst others on exact location in the machine and loading type.

However, the stationary residuals of this long term relation are not sufficient for monitoring the process and detecting the out of control points caused by the blocking of the bearings. If  $\mathbf{x}_t$  and  $\mathbf{y}_t$  are cointegrated an Error Correction Model (ECM) is useful for describing the short term dynamics between them. This ECM is given by

$$\Delta \mathbf{y}_t = \alpha + \beta \Delta \mathbf{x}_t + \gamma (\mathbf{y}_{t-1} - c_0 - c_1 \mathbf{x}_{t-1}) + \delta_t. \quad (13)$$

The parameter  $\gamma$  measures the speed of convergence of the two series towards the long term equilibrium,

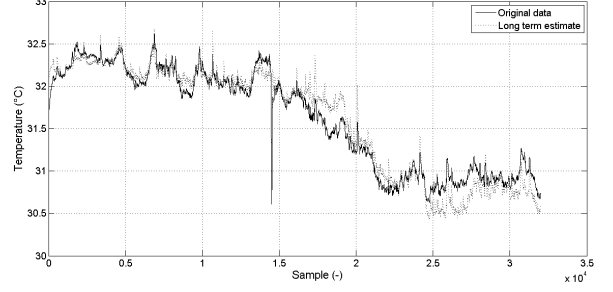


Figure 13: Long term relationship between the two temperature channels.

and should have a negative value to be interpretable (indeed, negative means that they are attracted towards the equilibrium).  $\alpha$ ,  $\beta$  and  $\gamma$  are parameters to be estimated and  $\delta_t$  is the error term which is assumed to be i.i.d. Fitting this ECM is typically performed in two steps. First,  $c_0$  and  $c_1$  are estimated using the long term relation of Eq. 11. Then, their estimates are plugged into Eq. 13 to yield  $\alpha$ ,  $\beta$  and  $\gamma$ . Fitting the ECM on the temperature dataset is shown in Figure 14 and results in an almost perfect fit of  $\mathbf{x}_t$  and  $\mathbf{y}_t$ , so that the residuals are expected to be highly informative. Formally checking those residuals leads to the conclusion that they are i.i.d. After fitting the ECM on the given dataset, it was tested on a separate dataset. The residuals for this validation set are shown in Figure 15. They are clearly stationary (ADF,  $p < 0.001$ ) and the fault states which occurred after 250, 1,200, 4,950, 5,700 and 5,900 samples are easily discernible using simple control charts.

### 3.3 Discussion and future perspectives

For the simple example of a bivariate, nonstationary series we used cointegration to turn the data into a simple, univariate, stationary series of residuals that is easy to monitor. As such, the example proves the potential of cointegration and

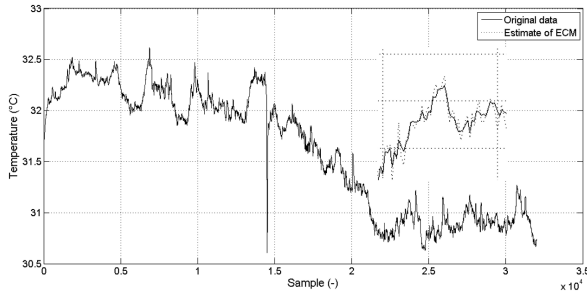


Figure 14: The Temperature dataset: short term relationship through the Error Correction Model. Zoomed view added to show the difference between the data and the estimates.

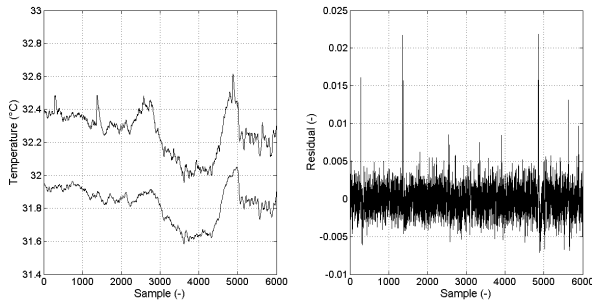


Figure 15: Two temperature series for validating the model (left). Residuals of the ECM for the validation set, with fault states clearly visible (right).

the related models to cope with nonstationarity. However, to turn this simple example into a robust solution for contemporary processes where the dimensionality is often substantially higher, additional research is required. We list here some extensions, issues and potential avenues for further research.

Although we used only the bivariate example, the cointegration framework extends well beyond this and is capable of modelling multivariate time series. In that case there is eventually more than one cointegration equation that needs to be estimated. Johansen (1991) developed a test to determine in a sequential way the number of cointe-

gration equations that describe the data (the Johansen test). When the multivariate series are cointegrated, the ECM we discussed before can be expanded to a Vector ECM (VECM), much like an AR process is expanded to a VAR model. In fact, the VECM is much like a VAR model for the series in differences (first derivative), but with the addition of the cointegration equations that are found significant using the Johansen test.

A potential issue arises when we want to apply the cointegration principle to multivariate processes where a substantial number of variables are measured. In such cases, the number of parameters is often prohibitive and can even be larger than the number of samples taken. This fact is further complicated if the cross-correlation between the observed variables is high so that the effective sample size is further reduced. This situation is becoming increasingly common, for instance in cases where spectral data (near infrared, mass spectrometry, vibration spectra, ...) are acquired. This situation inspired several researchers to develop adapted techniques, which are most often based on penalization techniques (e.g. lasso (Song and Bickel, 2011)). An appealing approach to tackle the dimensionality issue is the use of Principal Component Analysis in combination with cointegration. Peña and Poncela (2006) described the use of PCA in estimating VAR models and tested it successfully on a real-life dataset, although it only featured seven variables, providing a compelling case for an extension of VECM in this direction.

Another potential disadvantage of the cointegration assumptions is the fact that the parameters of the long term relationship dictated by the cointegration equation (Eq. 11) do not change over time, neither do the parameters of the ECM. If this assumption does not hold, the residuals of the cointegration approach will no longer be stationary and monitoring is impeded. Hansen and Johansen (1999) described tests for the evaluation of parameter constancy in cointegrated vector autoregressive models, and proposed two different ways of re-estimating the parameters of the model. This

approach seems to be relevant for industrial applications, but lacks the ability to cope with high-dimensional signals. This leads to the logical suggestion of combining the adaptive PCA models we have described above with the cointegration approach of this section.

## 4 Conclusions

In this paper we have highlighted the variety of challenges posed by time-dependent processes. In Section 2, we showed that monitoring high-dimensional processes with autocorrelation can be successfully achieved using PCA-based methods. However, for the ARI(1,1) processes, we found that even methods which are purportedly designed to address nonstationarity (RPCA and MWPCA) have difficulties in specifying a suitable model for that process type. Extensions to RPCA and MWPCA exist in the literature (for examples, see (De Ketelaere et al., 2015)) that may offer improvements over the basic implementations. However, these methods have not been thoroughly compared in the literature, so it is difficult to recommend one in particular.

In Section 3, we proposed that one potential step towards accurate modeling of nonstationary processes could come from approaches that make use of the concept of cointegration. When appropriate, these approaches may result in a more valid model of the processes than adaptive methods, like RPCA or MWPCA. However, classical methods for cointegrated data, such as the VECM, do not scale well to the high-dimensional setting. For this reason, further research on the integration of cointegration methodology into latent variable-based methods may be useful.

## References

- Bersimis, S., Psarakis, S., and Panaretos, J. (2007). "Multivariate statistical process control charts: an overview". *Quality and Reliability Engineering International*, 23(5), pp. 517–543.
- Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification". *The Annals of Mathematical Statistics*, 25(2), pp. 290–302.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall, New Jersey, 3rd edition.
- Burnham, A. J., MacGregor, J. F., and Viveros, R. (1999). "Latent variable multivariate regression modeling". *Chemometrics and Intelligent Laboratory Systems*, 48(2), pp. 167 – 180.
- Camacho, J., Pic, J., and Ferrer, A. (2009). "The best approaches in the on-line monitoring of batch processes based on PCA: Does the modelling structure matter?". *Analytica Chimica Acta*, 642(12), pp. 59 – 68.
- Chen, Q., Kruger, U., and Leung, A. Y. T. (2009). "Cointegration Testing Method for Monitoring Nonstationary Processes". *Industrial & Engineering Chemistry Research*, 48(7), pp. 3533–3543.
- Choi, S., Martin, E., Morris, A., and Lee, I.-B. (2006). "Adaptive Multivariate Statistical Process Control for Monitoring Time-Varying Processes". *Ind. Eng. Chem. Res*, 45, pp. 3108–3118.
- De Ketelaere, B., Hubert, M., and Schmitt, E. (2015). "Overview of PCA-based statistical process monitoring methods for time-dependent, high-dimensional data". *Journal of Quality Technology*.
- De Ketelaere, B., Mertens, K., Mathijs, F., Diaz, D., and Baerdemaeker, J. (2011). "Nonstationarity in statistical process control: issues, cases,

- ideas". *Applied Stochastic Models in Business and Industry*, 27(4), pp. 367–376.
- Engle, R. F. and Granger, C. W. J. (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing". *Econometrica*, 55(2), pp. 251–276.
- Ferrer, A. (2007). "Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process". *Quality Engineering*, 19(4), pp. 311–325.
- Ferrer, A. (2014). "Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift". *Quality Engineering*, 26(1), pp. 72–91.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York. John Wiley and Sons.
- Granger, C. and Newbold, P. (1974). "Spurious regressions in econometrics". *Journal of Econometrics*, 2, pp. 111–120.
- Hansen, H. and Johansen, S. (1999). "Some tests for parameter constancy in cointegrated VAR-models". *Econometrics Journal*, 2(2), pp. 306–333.
- Jackson, J. E. and Mudholkar, G. S. (1979). "Control Procedures for Residuals Associated With Principal Component Analysis". *Technometrics*, 21(3), pp. 341–349.
- Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models". *Econometrica*, 59(6), pp. 1551–1580.
- Jolliffe, I. (2002). *Principal Component Analysis*. Springer, New York, 2nd edition.
- Jyh-Cheng, J. (2010). "Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms". *Journal of the Taiwan Institute of Chemical Engineer*, 44, pp. 475–481.
- Kourti, T. (2005). "Application of latent variable methods to process control and multivariate statistical process control in industry". *International Journal of Adaptive Control and Signal Processing*, 19(4), pp. 213–246.
- Kruger, U. and Xie, L. (2012). *Advances in statistical monitoring of complex multivariate processes: with applications in industrial process control*. John Wiley and Sons Ltd.
- Kruger, U., Zhou, Y., and Irwin, G. W. (2004). "Improved principal component monitoring of large-scale processes". *Journal of Process Control*, 14(8), pp. 879–888.
- Ku, W., Storer, R. H., and Georgakis, C. (1995). "Disturbance detection and isolation by dynamic principal component analysis". *Chemometrics and Intelligent Laboratory Systems*, 30(1), pp. 179–196.
- Lakshminarayanan, S., Shah, S. L., and Nandakumar, K. (1997). "Modeling and Control of Multivariable Processes: Dynamic PLS Approach". *AIChE Journal*, 43(9), pp. 2307–2322.
- Li, W., Yue, H. H., Valle-Cervantes, S., and Qin, S. J. (2000). "Recursive PCA for adaptive process monitoring". *Journal of Process Control*, 10(5), pp. 471–486.
- Maddala, G. and Kim, I. (2003). *Unit Root, Cointegration and Structural Change*. Cambridge University Press, Fifth Edition, UK.
- Montgomery, D. (2008). *Introduction to Statistical Quality Control*. Wiley Desktop Editions Series. Wiley.
- Oppenheim, A. V. and Schaffer, R. W. (1975). *Digital signal processing*. Prentice-Hall, New York.
- Peña, D. and Poncela, P. (2006). *Dimension Reduction in Multivariate Time Series*, book section 28, pp. 433–458. Statistics for Industry and Technology. Birkhuser Boston.

- Phillips, P. (1986). "Understanding Spurious Regressions in Econometrics". *Journal of Econometrics*, 33, pp. 311–340.
- Phillips, P. (1987). "Time series regression with unit roots". *Econometrica*, 55, pp. 277–301.
- Ramaker, H.-J., van Sprang, E. N., Westerhuis, J. A., and Smilde, A. K. (2005). "Fault detection properties of global, local and time evolving models for batch process monitoring". *Journal of Process Control*, 15(7), pp. 799 – 805.
- Rato, T., Schmitt, E., De Ketelaere, B., Hubert, M., and Reis, M. (2015). "A Systematic Comparison of PCA-based Statistical Process Monitoring Methods for High-dimensional, Time-dependent Processes". *AIChE*, *accepted*.
- Rato, T. J. and Reis, M. S. (2013a). "Defining the structure of DPCA models and its impact on process monitoring and prediction activities". *Chemometrics and Intelligent Laboratory Systems*, 125, pp. 74–86.
- Rato, T. J. and Reis, M. S. (2013b). "Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR)". *Chemometrics and Intelligent Laboratory Systems*, 125, pp. 101–108.
- Schmitt, E., De Ketelaere, B., Rato, T., and Reis, M. (2014). "The Challenges of PCA-Based Statistical Process Monitoring: An Overview and Solutions". *ENBIS. Linz, Austria, 21-25 September 2014*.
- Song, S. and Bickel, P. J. (2011). "Large vector auto regressions". *Arxiv preprint arXiv:1106.3915v1*.
- Valle, S., Li, W., and Qin, S. J. (1999). "Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods". *Industrial & Engineering Chemistry Research*, 38(11), pp. 4389–4401.
- van Sprang, E. N., Ramaker, H.-J., Westerhuis, J. A., Gurden, S. P., and Smilde, A. K. (2002). "Critical evaluation of approaches for on-line batch process monitoring". *Chemical Engineering Science*, 57(18), pp. 3979 – 3991.
- Vanhatalo, E. and Kulachi, M. (2015). "The effect of autocorrelation on the Hotelling  $T^2$  control chart". *Quality and Reliability Engineering International*, 10.1002/qre.1717.
- Wang, X., Kruger, U., and Irwin, G. W. (2005). "Process Monitoring Approach Using Fast Moving Window PCA". *Industrial & Engineering Chemistry Research*, 44(15), pp. 5691–5702.
- Wikstrom, C., Albano, C., Eriksson, L., Johansson, E., Sandberg, M., Kettaneh-Wold, N., Fridén, H., Rännar, S., Nordahl, Å., and Wold, S. (1998). "Multivariate process and quality monitoring applied to an electrolysis process: Part II. Multivariate time-series analysis of lagged latent variables". *Chemometrics and Intelligent Laboratory Systems*, 42, pp. 233–240.
- Woodall, W. and Montgomery, D. (2014). "Some Current Directions in the Theory and Application of Statistical Process Monitoring". *Journal of Quality Technology*, 46(1), pp. 78–94.